# The Role of Cognitive Ability, Personality Traits and Gender in Gift Exchange Outcomes[1]

Emel Filiz-Ozbay
Department of Economics
University of Maryland

John C. Ham
Department of Economics and MPRC
University of Maryland
IFAU, IFS, IRP (UW) and IZA

John H. Kagel
Department of Economics
Ohio State University

Erkut Y. Ozbay
Department of Economics
University of Maryland

May 6, 2013

## Abstract

We examine the role of cognitive ability, personality traits and gender in a one-shot gift exchange experiment. Controlling for cognitive ability and personality characteristics, men offer higher wages than women do, as do agents with greater cognitive ability and those scoring higher on agreeableness on the Big Five personality scale. In turn, men provide greater effort than women do on average, and respond at higher wage rates with greater increases in effort. For both genders, a one standard deviation increase in agreeableness generates almost the same increase in effort as a comparable increase in wages. Omitting cognitive ability from the analysis, and pooling men and women, produces seriously biased parameter estimates.

Key words: gift exchange experiment, Big Five personality characteristics, SAT scores.

JEL classification: D03, J16, J22.

---

In economics, there is growing interest in the effects of personality traits, cognitive ability, gender, and ethnicity on economic outcomes using field data (Borghans et al., 2008; Heckman et al., 2006). In particular, cognitive ability, agreeability and conscientious have been found to be strongly associated with positive labor market returns (Urzua and Veramendi, 2012 and Kern et al., 2013). Labor market studies have the advantage that they examine high stakes behavior, but have the disadvantage that they necessarily aggregate over many different labor market situations and the effect of personality traits, in particular, may remain something of a black box. By studying the effect of cognitive ability and personality traits in a number of labor market environments in the lab, one hopefully will able to gain a much better handle on the ways in which cognitive ability and personality traits affect economic success.

We begin such a research program by looking at the effect of cognitive ability and personality traits on behavior in a one-shot laboratory gift exchange game. We start with the one-shot gift exchange game because it has been widely reported on and because both cognitive ability and noncognitive factors have a potentially important role to play here, given that agents with standard selfish preferences should play the income maximizing Nash equilibrium of zero wages and zero effort. The fact that subjects generally do not behave this way has important implications for the social preference literature as well as significant implications for a number of puzzling outcomes in labor economics from the viewpoint of standard economic theory (e.g., wage rigidity, rent sharing, and certain types of efficiency wages).[2] Our results also have implications for personnel economics (see e.g. Lazear and Oyer, 2013), since companies often test potential hires to learn about both their cognitive ability and personality traits in order to insure a better fit between the company and the potential employee. Finally, it is widely known that subjects exhibit substantial and persistent heterogeneity in both wage offers and effort responses in laboratory gift exchange experiments, so that it is interesting to ask how much accounting for cognitive ability and personality traits explains this unobserved heterogeneity.

---

[2] See Gächter and Fehr (2001) and Cooper and Kagel (2012a) for elaboration of these points.

In our analysis we focus on personality traits as described by the Big Five personality characteristics, and measured by the Big Five Inventory (BFI) – see John et al. (2008). The 44-item BFI was developed with the goal of creating a brief inventory that would allow robust, efficient and flexible assessment of personality without the need for more differentiated measurement of its individual elements, and has gained widespread acceptance in the psychology literature. It has also been used in a number of labor market studies when available. The properties of the BFI are discussed below, but for the moment, we simply note that the BFI provides measures of a person's agreeableness, extroversion, conscientiousness, neuroticism and openness.

We use SAT scores to measure cognitive ability, since they are measured with relatively little error (as demonstrated by similar test scores obtained by those taking the test multiple times) and exhibit relatively wide variation in our sample. However, we fully acknowledge that SAT scores are not a pure (fluid) intelligence measure and probably are affected by the Big Five characteristics. However, exactly the same issue arises with the Armed Services Vocational Aptitude Battery (AFQT) score used in the labor market studies discussed above, and the correlation between SAT and AFQT scores can be as high as 0.86.[3] Further, it is worth noting that one would also expect Big Five traits to be affected by cognitive ability; e.g., people with greater cognitive ability are likely to learn more quickly that agreeableness typically pays off in life. Moreover, SAT scores have a substantial positive correlation with a variety of traditional IQ measures (Frey and Detterman, 2004). Finally, SAT scores have the advantage of being readily available for our subject population.

We find substantial gift exchange with the usual pattern reported in the literature: higher wages result in substantially higher average effort levels, which are mutually profitable for both "managers" and "workers." For the pooled data we find that the coefficients on the gender dummy in wage offers and the gender by wage interaction in effort responses are significantly negative. We also run fully separate models for men and women since the data indicate that this

---

[3] The AFQT test is based on a subset (word knowledge, paragraph comprehension, arithmetic reasoning, and mathematics knowledge) of the ten-section ASVAB test that is administered to military recruits. The ASVAB test often is seen as an entrance exam for the military, although the army describes it differently: "the test is to measures your knowledge and ability in ten different areas. It is not an IQ test, but the ASVAB is one of the ways to help you decide what job areas in the Army would be best for you (see http://www.goarmy.com/learn/understanding-the-asvab.html). The AFQT test is not publically available so we cannot use it here. Previous researchers have used it in analyzing field data from National Longitudinal Survey as the NLS contains AFQT scores for all individuals in the sample.

is appropriate. We find that SAT and agreeableness have important effects for offered wages in the pooled estimates, as well as when we look at men and women separately. However, if we drop SAT scores from the pooled wage offer regression, there is no significant effect for agreeableness on wage offers, contradicting results from both field and experimental studies. This indicates the need to control for cognitive ability in assessing the impact of personality characteristics.[4] The importance of allowing for gender differences over the full set of personality characteristics is demonstrated by the fact that conscientiousness plays no role in the pooled data for wage offers, but does when looking at men and women's wage offers separately, as a consequence having a different sign for men and women. We check to see if these differences between men and women reflect unobservable differences in terms of who volunteered for the experiment using balancing tests drawn from the labor economics literature, and determine this is not the case.

The impact of personality characteristics on behavior can be quite substantial. For example, a one standard deviation increase in the SAT score is associated with an increase in the wage offer index function of men and women of 30.8% and 42.9% respectively. Further a one standard deviation increase in agreeableness is associated with an increase in the wage offer index function of men and women of 15.9% and 16.2% respectively, while a one standard deviation increase in conscientiousness *increases* the wage offer index by 23.4% for men but *lowers* it by 32.8% for women. Similar results hold for effort responses: A one standard deviation increase in wages increases the effort response index function by 135.4% and 100.0% for men and women respectively. SAT affects the effort response of men but not women, and for men a one standard deviation increase in SAT *decreases* the effort response index function by 44.9%. Further, a one standard deviation increase in agreeableness increases the effort response index function by 87.4% and 106.5% for men and women, respectively, an impact comparable to a one-standard deviation increase in wages. Finally, a one standard deviation increase in conscientiousness increases the effort response index function for men by 19.2% but lowers it for women by 6.5%.[5] Dropping SAT leads to a large increase in the wage coefficient in the effort response index function for men but not women, while dropping the Big 5 does not affect the remaining

---

[4] Significant agreeableness effects are replaced by statistically significant, negative effects of extroversion on wage offers, which has no precedent in the literature.
[5] One caveat here is that neither effect is statistically significant at the mean values.

coefficients for either men or women.[6] Thus, we are largely able to mimic the labor market effects of cognitive ability and personality traits; the exception is the different effect of conscientious on women's and men's wage offers.

Consistent with the gift exchange literature we also find a large amount of persistent unobserved heterogeneity across subjects. However, we find that adding the Big 5 and SAT scores reduces the variance of persistent individual unobserved heterogeneity by 37% and 23% in the wage offer and effort response pooled estimates, respectively, with similar results when looking at men and women separately. Thus, adding measures of cognitive ability and personality traits goes some way towards reducing the role of subject heterogeneity in gift exchange outcomes. Finally, it is interesting to note that the variance of persistent individual heterogeneity for men is 50% and 100% larger than that for women in the wage offer and effort response index functions respectively, a difference which has not been noted in the literature prior to this.

We briefly contrast these results with earlier ones from the experimental literature. Casari et al. (2007) found that both men and women with higher cognitive ability (as measured by SAT scores) were better able to avoid the worst effects of the winner's curse when bidding in common value auctions; they controlled for gender differences but not for personality traits in their analysis.[7] Kurzban and Houser (2001) looked at the role of the Big Five personality characteristics, along with other personality measures, on behavior in a voluntary contribution mechanism (VCM) public good game, using a multinomial logit model.[8] They found no statically significant relationship between contribution levels and any of the Big Five characteristics.[9] They attribute this finding, in part, to their relatively small sample size (57 subjects). Pothos et al. (2010) investigated individual correlations between the Big Five components on cooperation in a simultaneous move, one-shot prisoner's dilemma game and

---

[6] Neither the male or female wage coefficients are statistically significant with or without SAT, but dropping SAT certainly shifts the confidence interval upward for the male wage coefficient.

[7] We are confining this short revue of prior experimental work to those papers looking at the impact of the Big Five and other personality characteristics on social preferences, which our paper deals most directly with. As Borghans et al., (2008) note there has been limited research on personality characteristics on this important domain. For a survey regarding the power of personality traits both as predictors and as causes of academic and economic success, health, and criminal activity see Almlund et al. (2011).

[8] They classified people into three groups: (1) free riders, (2) cooperators, and (3) conditional cooperators.

[9] They report that men were more cooperative than women, while those with higher self-esteem (not one of the Big Five) tended to free ride more.

found that more agreeable types were less likely to defect (p = 0.054).[10] Becker et al. (2012) investigated the relationship between the Big Five personality characteristics and behavior in a variety of one-shot games with strong social preference elements - the trust game, investment in punishment in a public good (VCM) game, and giving in the dictator game. They focus on direct correlations between the Big Five personality measures and measures of the strength of social preferences in these games.[11] Among all the personality characteristics, agreeableness had the largest and most significant correlations in their study, being positively correlated with second mover returns and first mover allocations in the trust game, along with giving in the dictator game, and negatively correlated with punishment in the VCM game. Openness had the same qualitative (but weaker quantitative) relationship to social preferences in these games. None of these papers controlled for any measure of cognitive ability in their analysis, nor do they investigate the effect of gender differences over the full set of characteristics.

Anderson et al. (2011) is the paper closest to ours. Using a large sample of truck driver trainees, they measure individual risk and time preferences, along with obtaining scores for the Big Five personality characteristics and cognitive ability based on the Cognitive Skill Index.[12] They run regressions investigating the impact of these variables on a number of life outcomes (e.g., filing for bankruptcy, smoking, and credit scores), while also controlling for socio-economic characteristics (e.g., age, education and marital status). Their overlap with our paper comes when they turn their attention to behavior in a modified (one-shot) trust game in which first movers could send either $0 or $5, and second movers responded via the strategy method.[13] To the best of our knowledge, they are the only other experimental study to control simultaneously for personality characteristics and cognitive ability in a laboratory type setting. They found that more agreeable types and those scoring higher on cognitive skills were more likely to send the $5, while more conscientious types were less likely to do so. More agreeable and more neurotic types were more likely to send money back in response to either a $0 or $5

---

[10] They employed procedures very different from the typical economic experiment, using loaded terms (e.g. "defect" or "cooperate") with no financial incentives. Jones (2012) reports similar results for a random end point, repeated PD game using standard experimental economic procedures.

[11] These direct correlations do not control for any of the other Big Five characteristics. They also check for potential non-linear relations, which would compromise the correlation analysis, but report minimal nonlinearities.

[12] This index is computed as a function of: i) the first factor in the factor analysis of the Raven's score; ii) the Numeracy score; and iii) the score in the Hit 15 game, and appears to be original to their paper.

[13] Under the strategy method, second movers state their actions contingent on first mover behavior and then are bound by these actions.

transfer, with higher cognitive ability types sending less (more) back in response to a $0 ($5) transfer. In addition, women were less likely than men ($p < 0.10$) to send money back in response to receiving $5.[14]

Our paper complements their analysis of social preferences in several ways, which is important given the need for a wide-range of studies on these issues. First, we investigate the effect of cognitive ability and personality characteristics separately for men and women, finding a number of important differences. Second, we use a different measure of cognitive ability, and do not use a finer breakdown of personality traits than that given by the Big 5, as suggested by the psychology literature. Third, we consider the biases in estimated Big Five effects that occur when SAT is omitted, and the potential biases in estimated SAT effects that occur when the Big 5 are omitted. Fourth, we use a sample that is much more comparable to those used by previous experimenters, and this allows us to suggest that previous results on personality characteristics are likely to be misleading since they did not control for cognitive ability and a full set of gender differences. Finally, we examine how much of subject specific heterogeneity in gift exchange experiments can be controlled for by using SAT and Big 5 scores as conditioning variables.

The structure of the paper is as follows: Section 1 provides a characterization of the Big Five personality characteristics, focusing on what they are designed to identify. Here we also discuss the pros and cons of using SAT scores to measure cognitive ability. Section 2 discusses the experimental design and procedures, along with the anticipated effects of cognitive ability and the Big Five on wage offers and effort responses. Experimental results are reported in Section 3, starting with the usual measures reported for gift exchange experiments, and then moving on to the main analysis regarding the effects of gender, cognitive ability and the Big Five characteristics on outcomes reported. Section 4 summarizes the main results of the paper and suggests additional areas of study.

### 1. Big Five Personality Measures and SAT Scores

*1.1 Choice of Measures of Cognitive Ability and Personality Traits*

Prior to the start of each session, subjects filled out the 44-item Big Five Inventory (BFI) questionnaire (John et al., 2008).[15] The Big Five personality characteristics represent a consensus

---

[14] A large portion of their sample of truck drivers were unconditional cooperators, sending money back regardless of whether or not they received anything from senders.

among personality psychologists on a general taxonomy of personality traits. These personality characteristics do not represent a particular theoretical perspective but are derived from the analysis of the natural language terms people use to describe themselves and others. The focus of the Big Five is on internal consistency rather than predictive ability. The idea behind the Big Five is not that these personality characteristics reflect any intrinsic greatness or that personality differences can be reduced to five traits but that the five dimensions represent personality at a very broad level of abstraction; each dimension summarizes a large number of distinct, more specific, personality characteristics. When more factors than the Big Five have been identified across cultures and studies, they rarely replicate across multiple studies conducted by independent investigators.

The BFI measure consists of 44 short phrases based on trait adjectives known to be prototypical markers for the Big Five. For example, the openness adjective *original* becomes "Is original, comes up with new ideas" in the BFI. There are between eight and ten phrases associated with each of the five subscales, which are interspersed with each other using both reverse-keyed and direct-keyed items, designed to disguise the nature of the characteristics in question.[16] Further, in a study such as this one where one wants to relate the BFI measures to behavior, subjects are not told either the purpose of the questionnaire or their scores on it.

A number of different instruments are available to measure the Big Five, none of which is considered to stand out as the gold standard. The BFI has been used frequently in research settings where time is at a premium, as it typically takes between 10 and 15 minutes to complete. The Big Five personality traits consist of:

(1) Agreeableness – contrasts a pro-social and communal orientation to others, and includes traits such as altruism, tender-mindedness, trust and modesty.

(2) Extroversion – implies an energetic approach toward the social and material world, including traits such as sociability, activity, assertiveness and positive emotionality.

---

[15] These were computerized using zTree (Fischbacher, 2007), the software used to program the experiment. The BFI questionnaire is available at http://www.outofservice.com/bigfive/. The material in this section on the Big Five is taken from John et al. (2008) and Borghans et al. (2008).

[16] For example, two of the agreeableness subscales questions consist of question (2) "Tends to find fault with others" which is reverse-keyed so that a higher rating on this scale counts as less agreeable, and question number (7) "Is helpful and unselfish with others" where a higher rating counts as more agreeable.

(3) Conscientiousness – describes socially prescribed impulse control that facilitates task- and goal-directed behavior, such as thinking before acting, delaying gratification, following norms and rules, and planning, organizing and prioritizing tasks.

(4) Neuroticism – contrasts emotional stability and even-temperedness with negative emotionality, such as feeling anxious, nervous, sad and tense.

(5) Openness – describes the breadth, depth and complexity of an individual's mental and experiential life.

Scoring higher on the scale of each characteristic is associated with the more positive elements of the traits, except for neuroticism, where the high pole is associated with poorer coping with life events.

As noted above, we use SAT scores as our proxy measure for cognitive ability (denoted by $g$) because they are readily available through the registrar's office (with subjects having signed consent forms to this effect), are measured with relatively little error, and exhibit relatively wide variation in our sample.[17] They are highly correlated with the measure, AFQT score, that is used in labor market field studies, and relatively highly correlated with scores from Raven's Advanced Progressive Matrices (hereafter Raven's measure).[18] Like AFQT scores, SAT scores can be affected by learning (and thus by the Big Five characteristics), as well as by cultural factors. Interestingly, research in the psychology literature indicates that Raven's measures also are affected by cultural background and education level (Brouwers et al., 2009), instruction in taking the test (Skuya et al., 2002), and practice in taking the test (Bors and Vigneaub, 2001). Moreover Raven's measures tend to exhibit high variance for the same individual (Bors and Vigneaub, 2001), suggesting that they contain more measurement error than SAT scores; such measurement error is likely to bias both the cognitive ability and Big 5 coefficients if the latter are correlated, as is the case in our data. Thus, we conclude that there appears to be no dominant measure of cognitive ability for laboratory experiments, so that a study comparing results using different measures of cognitive ability would be very useful.

---

[17] We use the overall score, as Casari et al. (2007) found that this was better at predicting behavior than the math or verbal score alone.

[18] For example, Frey and Detterman (2004) found that the correlation between SAT scores and the Raven's measures among 104 undergraduates was 0.48 (0.72 for a restricted range).

*1.2 Summary Statistics*

The Big Five personality characteristics are on a scale of 1 to 5, with SAT on a scale of 400-1600. In the data analysis that follows, we convert these scores to the percent of maximum possible score (POMP). Specifically for individual $i$

$$\text{POMP}_i = \frac{Observed_i - Minimum}{Maximum - Minimum}$$

where *Observed$_i$* is the observed score for individual $i$, $Minimum$ is the minimum possible score on the scale, and $Maximum$ is the maximum possible score on the scale. POMP scores have a number of desirable characteristics for indexes of this sort, particularly in terms of the regression coefficient estimates that follow, as it puts them on a normalized scale that helps in interpreting the results (Cohen et al., 1999). Since POMP is a linear transformation of the original scores, statistical evaluation of the data remains unchanged. Table 1 reports average POMP scores, along with their ranges and standard deviations.

Table 1
POMP Scores for the Big Five and SAT

| Group | N | Statistic | SAT | Ope | Con | Ext | Agr | Neu |
|---|---|---|---|---|---|---|---|---|
| | | | | | | Scale | | |
| Total Population | 192 | Mean | 72.7 | 67.7 | 66.3 | 59.1 | 67.4 | 41.2 |
| | | SD | 11.8 | 15.0 | 13.9 | 18.6 | 16.1 | 18.9 |
| | | Range | 44 - 98 | 22 -100 | 25 – 98 | 12 - 97 | 13 - 100 | 3 – 94 |
| Males | 105 | Mean | 74.0 | 67.6 | 65.1 | 58.2 | 68.2 | 38.4 |
| | | SD | 12.4 | 14.6 | 14.2 | 17.7 | 16.5 | 18.5 |
| | | Range | 44 - 98 | 22 – 100 | 25 – 95 | 21 - 97 | 25 - 100 | 3 – 94 |
| Females | 87 | Mean | 71.1 | 67.8 | 67.8 | 60.2 | 66.4 | 44.5 |
| | | SD | 10.9 | 15.5 | 13.5 | 19.5 | 15.6 | 18.9 |
| | | Range | 45 - 97 | 25 – 90 | 30 – 98 | 12 - 97 | 13 - 95 | 9 – 91 |
| P values[a] | | | 0.087 | 0.892 | 0.185 | 0.455 | 0.459 | 0.027 |

Ope = Openness; Con = Conscientiousness; Ext = Extroversion; Agr = Agreeableness; Neu = Neuroticism.
[a]P values for equality of mean values between men and women.

## 2. Experimental Design and Procedures

After the completion of the BFI questionnaire, subjects were given written instructions, which were also read aloud.[19] After reading the instructions, subjects were randomly divided into two equal size groups: "managers" and "employees." Subjects played the same role throughout an experimental session. Each experimental session had 12 periods, with the number of periods announced in advance. At the beginning of each period, each manager was randomly matched with an (anonymous) employee to play the two-stage gift exchange game. There were 16 subjects in each session, with the random matching protocol programmed so that no employee was re-matched with the same manager more than twice and never re-matched in two consecutive periods. The anonymity, in conjunction with subjects learning the outcome of only their own match, was designed to generate a sequence of one-shot games.

In stage 1 of each period, managers chose a wage, which had to be an integer value from the interval [0,100]. In stage 2, each employee, after seeing the wage offer, chose an effort level, which also had to be an integer value from the interval [0,100], after which each manager observed the effort level of the employee he or she was paired with.[20] Payoffs were symmetric and calculated as follows for managers ($\pi_M$) and employees ($\pi_E$):

$$\pi_M = 100 - w + 5e$$
$$\pi_E = 100 - e + 5w$$

where *w* is the wage offered and *e* is the effort level chosen. Payoff functions were common knowledge, with subjects asked to calculate the payoffs of managers and employees in five examples before the experiment started, with the goal of ensuring that subjects understood the payoffs for themselves and the player they were paired with.

Assuming that players care only about their own income, the unique subgame perfect Nash equilibrium of the game is to provide zero effort after any wage offer, in anticipation of which wage offers are zero as well. On the other hand, the efficient wage and effort level, which maximize total surplus, is 100 for both. However, a large number of experiments with different

---

[19] The instructions of the experiment are available at www.econ.umd.edu/~filizozbay/FHKO_instructions.pdf.

[20] In addition to providing payoff information for the current period, each subject's computer screen reported the whole history of that subject's play and payoffs.

subject populations show that this outcome rarely occurs, with higher wage offers met with higher average effort responses.

Twelve sessions were run at the Experimental Economics Laboratory at University of Maryland. All sessions were computerized using the zTree software (Fischbacher, 2007). Subjects were recruited from the undergraduate population of University of Maryland using an online recruitment system. Sessions lasted about 60 minutes, including answering the BFI questionnaire. Subjects were paid privately and individually at the end of each session at a rate of 250 experimental currency units (ECUs) to 1 US dollar along with a $6 participation fee. Average earnings were approximately $21.75 for employees and $14.40 for managers (including the $6 show-up fee).

*2.1. Expectations Regarding Impact of Cognitive Ability, Big Five Personality Characteristics and Gender*

As noted above, cognitive ability, agreeability and conscientiousness have been most closely linked to positive labor market outcomes using field data. In what follows we discuss the potential impact of these factors as applied to the one-shot gift exchange game. Consider first cognitive ability, which under reasonable assumptions could have a negative or positive effect on wage offers. On the negative side, those with higher cognitive ability might be expected to offer lower wages, anticipating zero or low effort levels on account of the one-shot nature of the game. On the positive side, past experiments show that higher wage offers typically result in higher effort levels and greater profits for managers, which higher cognitive types might be better able to anticipate, or to respond to after some exploration. Further, since the extant literature indicates higher cognitive ability types tend to be less risk averse (Dohmen et al., 2010; Burks et al., 2009), managers with greater cognitive ability might also be more willing to bear the risk inherent in offering higher wages given the positive frequency of low effort levels reported at all wage rates in the typical gift exchange game. There are potential positive and negative effects of greater cognitive ability with respect to effort levels as well. Again, the one-shot nature of the game with its implications for lower or zero effort should be more transparent to workers with greater cognitive ability. On the other hand, workers with greater cognitive ability are more likely to be sensitive to the social norms of reciprocity in work relations. An additional motivating factor for positive reciprocity might be the desire to establish a good group reputation

effect (Kandori, 1992), the possibilities of which higher cognitive types might be more cognizant of. This possibility is, however, mitigated here by the fact that managers and workers only get to see their own outcomes, and that past experiments suggest that such groups reputation effects are hard to establish in groups of this size (e.g., Duffy and Ochs, 2009). At any rate, the group reputation motivation can be easily checked as it implies a rather sharp reduction in effort levels as the end stage draws near.

One unambiguous effect anticipated here is more agreeable types will offer higher wages and provide greater effort levels at each wage rate. This follows directly from the agreeability characteristic which contrasts a pro-social and communal orientation towards others with antagonism and includes traits such as altruism and trust. This is consistent with results from past experiments which show that in the trust game more agreeable types transfer more and return more (Becker et al., 2012)), and are more cooperative in repeated prisoner dilemma games (Jones, 2012). The real question here is what will be the relative impact of greater agreeability on outcomes, especially compared to the impact of increased wages on effort levels. As hard-nosed economists we might anticipate a substantially weaker agreeability effect on effort than wages, but some of our friends in psychology and management are not quite so sure about this.

Conscientiousness describes individuals who tend to follow norms and rules, with higher scorers having better job performance, the implications of which would appear to be somewhat ambiguous with respect to wage offers and effort levels in the gift exchange game. We might conjecture that, to the extent higher effort in response to higher wages is the accepted social norm, more conscientious types would be more likely to respond that way. The impact, if any, of the remaining Big Five personality characteristics have for the one-shot gift exchange game remain to be seen.

Although the gift exchange game has been studied with different sub-populations, to the best of our knowledge there have been no reports of gender differences in previous gift exchange experiments. Gender differences in trust and reciprocity, the characteristics most directly related to our experiment, have mostly been studied in the context of the trust game. A number of trust game experiments find no gender differences in sending behavior. However, other studies find that men are more trusting than women are. While some studies have found no gender differences in reciprocity, others have found that women are more reciprocal than men (see Croson and Gneezy, 2009, for a broad review of gender effects across a variety of economic

settings including the trust game). Risk preferences play a role for first movers in the trust game, as well as for wage setters in the gift exchange game. So that while women are generally reported to be more risk averse than men, Eckel and Wilson (2004) find no significant difference between risk measures and decisions to trust. Taken as a whole this would lead us to predict no gender differences in the gift exchange game, or that men will tend to offer higher wages than women as they tend to be more risk tolerant and to send more in the trust game, and that women will provide greater effort levels than men.

One concern when considering gender differences in an experiment is that the differences may simply reflect that the men and women recruited into the experiment are very different. Since we condition on SAT and the Big 5 in the statistical analysis, selection in terms of these variables will not be an issue, but men and women being different in terms of unobservables is a potentially confounding factor. To investigate this issue we use the labor economics approach of checking for differences between men and women in terms of observables, as a proxy for their differences in terms of unobservables. The University of Maryland is 53% male and 47% female, while our sample is 54.7% male and 45.3% female. In terms of the characteristics in Table 1, only the difference in neuroticism is significantly different at the 5% level. Since obtaining at least one significant coefficient at the 5% level could happen by chance 26.5% of the time even if the male and female values are equal, we do not reject the null hypothesis that the values in Table 1 are the same for men and women.[21] Further, we have data on majors for 50 men and 42 women of which 48% of the men and 38% of women are economics and business majors, 20% of the men and 19% of the women are science and engineering majors, and 32% of men and 43% of the women have "other" majors. None of these differences are significant at anything approaching standard confidence levels. Thus, we conclude that the men and women in our sample are very similar in terms of observed characteristics, so that it is reasonable to assume that they are not different in terms of unobservables.[22]

---

[21] This assumes independent draws, i.e. $0.265 = 1 - (0.95)^6$. We also have two significant differences at the 10% level. The probability of obtaining at least 2 differences at the 10% level when means are actually equal is 0.29.

[22] Note that we are not worrying about how our sample of men and women compare to all undergraduates at the University of Maryland.

## 3. Experimental Results

Aggregate outcomes are reported first in the way they are commonly reported for experiments of this sort. More detailed analysis follows, accounting for the impact of the BFI personality traits and SAT scores on behavior. Results are summarized in terms of a series of conclusions following the data reported.

### 3.1 Overview of Experimental Results

Figure 1 shows average wage offers and effort levels over time for the pooled data. Average wage offers were 50.7 ECUs with an average effort of 25.2 ECUs. As is typically the case, average wage offers and effort levels are relatively constant over time, with at most a very small (or nonexistent) end game effect with respect to effort levels.

Figure 2 shows effort as a function of the wage rate offered, with average effort clearly increasing with increases in the wage rate.[23] Although there was some heterogeneity between sessions, similar patterns are reported in all sessions. Observe that average wage offers are persistently higher than average effort levels, with neither of them close to the strictly selfish, own income maximizing, equilibrium of zero wage and zero effort. This was true for all sessions with only 12.0% of all wage offers and 37.3% of all effort choices equal to zero for the pooled data. Note that the vast majority of zero wage offers (92.0%) were met with zero effort. Excluding these zero wage offers, 29.9% of all effort choices were zero for the pooled data.



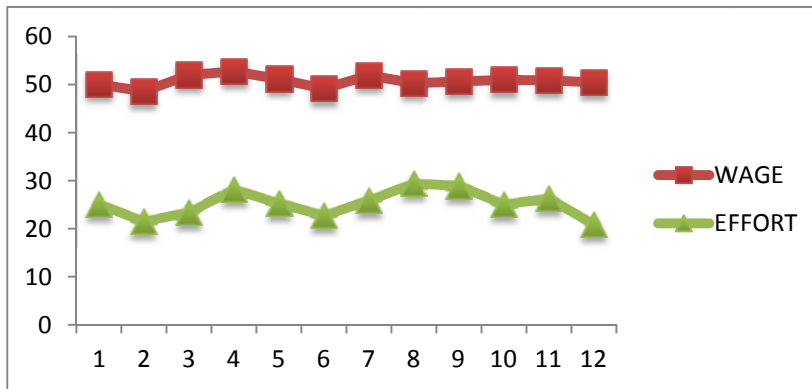**Figure 1.** Average wage and effort level per period.

---

[23] Following the standard for reporting in this literature, we include zero wage offers but delete them when analyzing responses using the Tobit analysis of effort response in section 3.3.
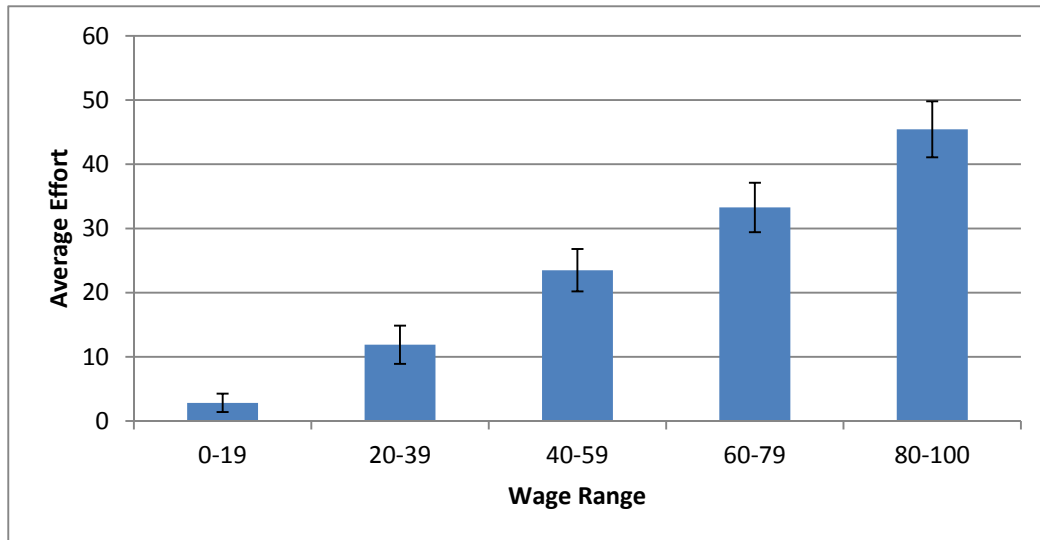
**Figure 2.** Effort level over each wage range for aggregate data - bars indicate 95% confidence interval for the mean.

Conclusion 1: *Aggregate wage and effort level data exhibit the typical pattern reported for one-shot gift exchange games: Average wages are higher than average effort levels, with more effort provided in response to higher wages. Further, the data show minimal, or non-existent, end game effects.*

Figure 3 reports average managers' payoffs in terms of the wage rate offered, net of the base rate of 100 ECUs in the managers' payoff function $\pi_M$ (where error bars indicate the 95% confidence interval for mean payoffs). As is typically the case, managers who offer higher wages are rewarded with substantially higher average incomes because of the substantially higher effort levels workers provide. However, as Figure 4 shows, these higher incomes are not without risk, as 20% or more of the effort responses at higher wage rates are met with zero effort.
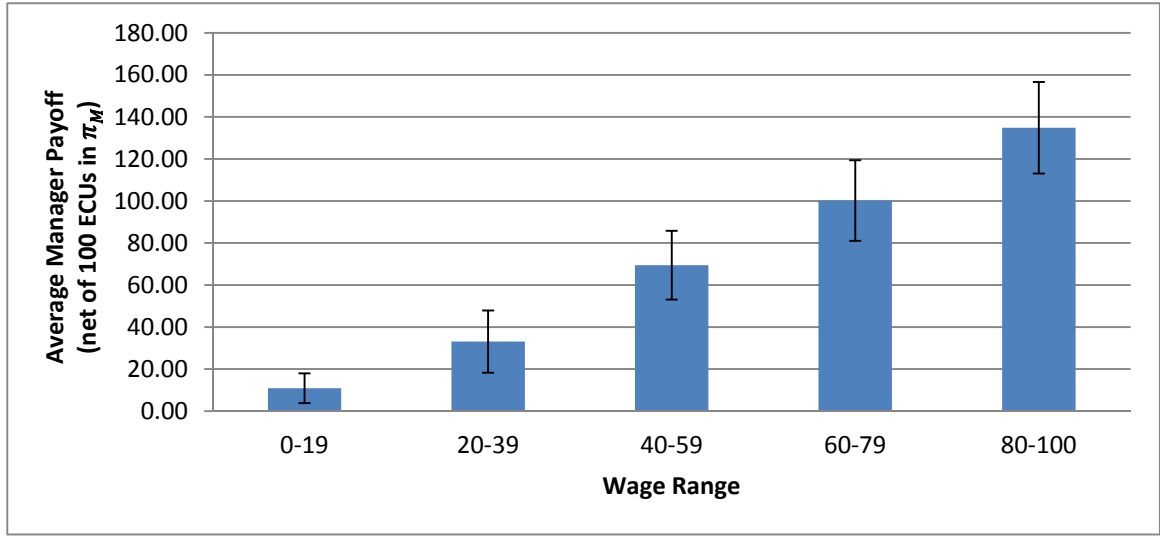
**Figure 3:** Average income of managers at each wage interval (net of the 100 ECUs included in $\pi_M$).
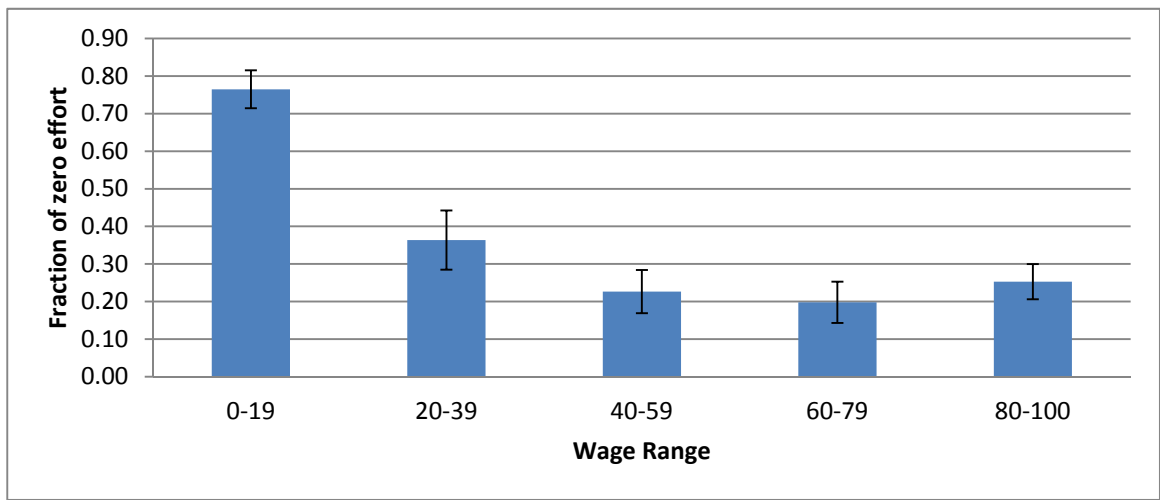


**Figure 4:** Fraction of zero effort at each wage interval.

Conclusion 2: *Offering higher wages yields greater expected income for managers. However, there are risks involved because of a persistent percentage of zero effort responses regardless of the wage rate offered.*

These typical aggregate level effects reported here for wage and effort levels provides assurance that our request for SAT scores and demographic data, along with answers to the BFI questionnaire had no impact on subjects' behavior.

Figure 5 shows that although there are minimal differences in effort responses between men and women at lower wages, at middle and higher wage rates men consistently provide

greater average effort than women do. Figure 6 shows that men also tend to offer higher wages than women with most of this tendency accounted for because of men providing wages offers in the interval 80-100 substantially more often than women: 39.8% of all men's wage offers versus 16.9% for women (see Figure 6).


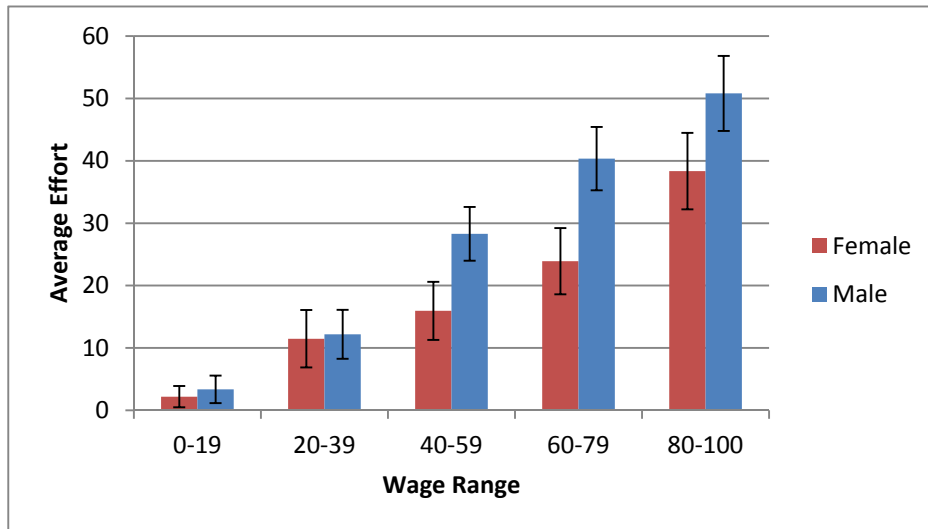
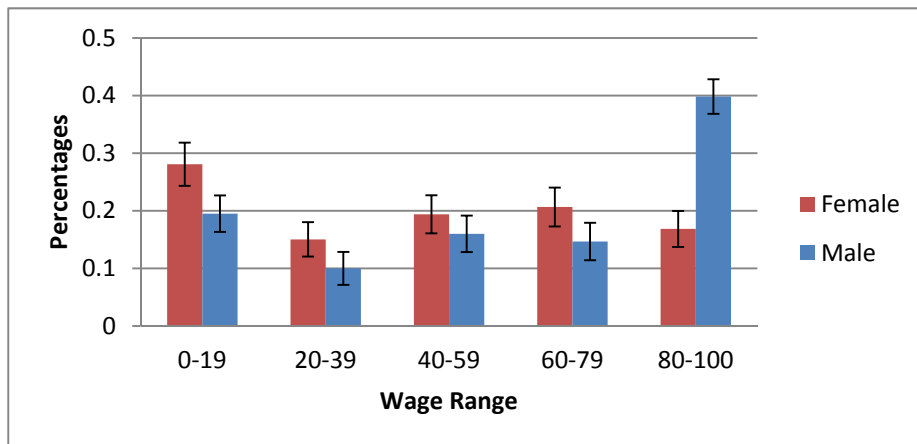**Figure 5.** Effort level over wage intervals for men and women.



**Figure 6**. Distribution of wage offers for men and women.

*Conclusion 3*: *Men tend to offer higher wages, and to have a bigger positive effort to higher wages, than women.*

Firm statistical support for Conclusion 3, along with the impact of the BFI and SAT scores on wage offers and effort responses, is provided in the regression analysis reported on below.

*3.2 Statistical Analysis Including the Big Five and SAT: Wage Offers*

Table 2 reports regression results for wage offers in relationship to subjects' SAT scores and the BFI.  Since actual wage offers are bounded by zero from below and 100 from above, a random effect, two-limit Tobit model is used for the statistical analysis. Specifically, we assume that the index function for offered wages for individual $i$ in period $p$ is

$$w_{ip}^* = \beta X_{ip} + \alpha_i + e_{ip} = \beta X_{ip} + u_{ip}. \tag{1}$$

Further, we assume that observed wage offers are determined by

$$
\begin{aligned}
w_{ip} &= 0 \quad \text{if} \quad w_{ip}^* < 0, \\
w_{ip} &= 100 \quad \text{if} \quad w_{ip}^* > 100, \\
w_{ip} &= w_{ip}^* \qquad \text{otherwise.}
\end{aligned}
\tag{2}
$$

In (1) $\alpha_i$ is a random effect error term, which is *iid* across $i$ and distributed as $N(0,\sigma_\alpha^2)$, while $e_{ip}$ is an idiosyncratic error term, which is *iid* over $i$ and $p$ and distributed as $N(0,\sigma_e^2)$.[24] The variance of $\alpha_i$, $\sigma_\alpha^2$, represents the persistent unobserved subject heterogeneity in wage offers.

We also considered a version of (1) that contained session fixed effects, but this did not have any effect on the standard errors, and the session effects were not close to being significant at standard levels.[25] All of the specifications included dummy variables for race – Caucasian, African-American, and Asian – with Other serving as the excluded category.[26] In almost all cases the coefficients on these variables failed to be statistically significant and are suppressed due to

---

[24] We also assume that $\alpha_i$ and $e_{jp}$ are independent for all $i$, $j$ and $p$, Thus the variance of the $u_{ip}$ is $\sigma_\alpha^2 + \sigma_\varepsilon^2$.

[25] We use session fixed effects instead of session random effects since the number of sessions is small (Imbens and Wooldridge, 2009).

[26] Other consisted of sixteen subjects: 7 Hispanics and 9 of unknown ethnicity.

space considerations. Those few cases where these variables proved to be statistically significant are discussed below in the text or in related footnotes.

It is worth emphasizing that $\beta_k$ represents the effect of increasing $X_k$ on the index function $w^*$ and not on the expected value of the observed wage $E(w) \equiv E(w \mid X)$. However, $\beta_k \geq 0$ implies that $\partial E(w)/\partial X_k \geq 0$ and vice-versa. Moreover, if $\beta_k$ is statistically significant, then $\partial E(w)/\partial X_k$ will also be significant. Finally, the larger $\beta_k$ is in absolute value, the larger $\partial E(w)/\partial X_k$ will be in absolute value as long as $E(w)$ is in the open interval $(0,100)$.[27] In what follows we will use 'the effect on the wage' and the 'effect on the offered wage index function' interchangeably for qualitative results, but will use the latter when reporting quantitative results.[28]

Columns 1 and 2 of Table 2 report estimates of the wage index function for the pooled data with and without the SAT variable respectively. Results absent SAT scores are reported because past studies have often looked at the impact of the Big Five while not having information on cognitive ability, and thus, by the standard Theil-Griliches specification error result, will produce biased coefficients for (1) if any of the Big 5 variables has a nonzero partial correlation with SAT.[29] Comparing the results with and without SAT indicates the magnitude of these biases, if any. The remaining columns repeat the analysis in columns (1) and (2) separately for men and women, since a likelihood ratio test for differences in coefficient values between the two rejects the null hypothesis that men and women have the same coefficients (p < 0.01) conditional on them having different intercepts.[30] This rejection is not surprising given the differences between men and women reported in Figure 6.

For the pooled data, SAT, agreeableness and the dummy variable for women are all statistically significant at conventional levels. Subjects with higher SAT scores offer higher wages, as do more agreeable types. Further, a likelihood ratio test rejects the null hypothesis that

---

[27] The difference between $\beta_k$ and $\partial E(w)/\partial X_k$ will be smaller here than in the standard Tobit model, since we have truncation at both ends.

[28] Labor economists refer to the terms $\partial E(w)/\partial X_k$ as 'partial effects' and often discuss these effects quantitatively. Here we focus on the Tobit coefficients because this approach follows the convention in experimental economics.

[29] Strictly speaking, this result applies to regressions and not Tobit results, but it is a helpful way to understand our results. Since the Theil-Griliches result will be unfamiliar to some experimental economists, in the Appendix we briefly review it and present the results of a regressing of SAT on the other explanatory variables for the pooled, male and female samples.

[30] The likelihood ratio test statistic equals 50.4, which is substantially larger than the critical value for $\chi^2_{99\%}(10)$.

the Big 5 variables jointly have zero coefficients.[31] The strong positive relationship between SAT and wages sorts out between competing possibilities in favor of the notion that managers with higher cognitive ability are better attuned to large potential profits associated with higher wages (recall Figure 3), and better able to tolerate the risk associated with offering these higher wages.

Table 2

Random Effects Estimates of the Wage Index Function

| Variable | Pooled data (men and women) | | Men | Women | Men | Women |
|---|---|---|---|---|---|---|
| SAT | 1.697*** (0.291) | | 1.530*** (0.360) | 1.667*** (0.464) | | |
| Agreeableness | 0.529** (0.206) | 0.372 (0.239) | 0.648** (0.280) | 0.453* (0.255) | 0.461 (0.324) | 0.300 (0.284) |
| Openness | -0.052 (0.196) | 0.058 (0.227) | -0.046 (0.296) | 0.157 (0.223) | 0.061 (0.347) | 0.228 (0.251) |
| Neuroticism | 0.084 (0.181) | 0.002 (0.210) | -0.093 (0.265) | 0.190 (0.206) | -0.115 (0.312) | 0.085 (0.231) |
| Conscientiousness | 0.191 (0.245) | 0.315 (0.284) | 1.147*** (0.355) | -1.042*** (0.339) | 1.477*** (0.406) | -1.23*** (0.378) |
| Extroversion | -0.102 (0.194) | -0.563*** (0.205) | -0.032 (0.272) | 0.343 (0.247) | -0.414 (0.298) | -0.012 (0.256) |
| Female | -15.23** (6.132) | -16.49** (7.149) | | | | |
| Period | -0.0906 (0.240) | -0.0902 (0.240) | 0.0035 (0.386) | -0.143 (0.295) | 0.007 (0.387) | -0.140 (0.294) |
| Constant | -97.73** (38.05) | 45.23 (33.78) | -142.7*** (50.53) | -91.40* (54.60) | -29.23 (49.96) | 63.11* (38.03) |
| Log-Likelihood | -4406.7 | -4421.4 | -2167.0 | -2214.5 | -2174.7 | -2220.2 |

*Statistically significant at the 10% level. ** Statistically significant at the 5% level. *** Statistically significant at the 1% level.

---

[31] The agreeableness outcome is not surprising in light of past research especially with respect to the trust game (e.g., Becker et al., 2012) and what the agreeableness characteristic is designed to represent.

The negative coefficient value for the gender dummy is consistent with findings that men are more trusting than women and/or more risk tolerant.[32]  Some may argue that these results provide a motivation for adding risk preferences or measures of trust and altruism as independent variables in the analysis. However, risk preferences are at least in part a function of cognitive ability and personality traits, and altruism and trust are components of agreeableness, so that including these measures would result in obtaining conditional estimates that understate the parameters of primary interest.[33] A second argument against breaking out the underlying components of agreeableness is that the psychology literature indicates that using a finer breakdown than the Big Five will lead to unstable parameter estimates across different applications.

Dropping SAT from the Tobits for the pooled data eliminates the statistical significance of agreeableness. Further, the extroversion characteristic, which is negative and not close to being statistically significant with SAT included, stays negative, increases substantially in absolute value, and becomes significant at the 0.01 level with SAT excluded. Using the specification error results in the Appendix for intuition, this change in value for extroversion reflects the fact that extroversion has a statistically significant negative coefficient in a regression of SAT on all of the other independent variables.[34]  Moreover, omitting SAT has a substantial effect on the intercept, as it goes from statistically significant and negative to positive but insignificant. However, dropping the Big 5 variables does not affect the SAT coefficient. Finally, adding the Big 5 and SAT scores to the probits reduces the variance of the persistent individual unobserved heterogeneity ($\sigma_\alpha^2$.) by 37% in the wage offer index function.

To obtain an idea of the magnitude of the effects implied by the estimates in column (1) of Table 2, Table 3 presents the effect on the wage index function of a one standard deviation increase in SAT and agreeableness (These are the variables that are statistically significant in

---

[32] For example,  it is far riskier to provide a wage offer in the interval of 80-100 compared to a zero wage offer with its guaranteed return of 100, the expected value of a wage offer in this interval (235 ECUs) far exceeds the guaranteed 100 ECUs for a zero wage offer.  Similarly, small, nonzero effort responses also make it risky to offer high wages.

[33] Comparable issues arise in working with field data where one has access to variables such as marriage and educational levels, since there is little doubt that the Big Five have an impact on these variables, so that to include them along with the Big Five is likely to mask the impact of the Big Five.  There is also the question of what measure of risk preferences to include since it is clear that rank order correlations of individual risk preferences across different domains, while not zero, are far from 1.0 (Einav et al., 2012).

[34] Its coefficient value is -0.142 (p < 0.01).

column (1); we also include conscientiousness since it becomes important once we separate men and women.) A one standard deviation increase in the SAT score increases the pooled wage index function by 19.8ECUs (38.4%), while a one standard deviation increase in agreeableness only increases the wage index function by 8.0 ECUs (15.5%).

Table 3

The Effect on the Wage Index Function of a One Standard

Deviation Increase in Key Explanatory Variables

(Change as a Percent of the Mean Value of the Wage Index Function in Parentheses)

| | | Change in Mean Wage Index Function Value Resulting from a One Standard Deviation Increase | | |
|---|---|---|---|---|
| | Mean Wage (*w\**) Index Function Value | SAT | Agreeableness | Conscientiousness |
| Pooled Data | 51.6*** | 19.8*** (38.4%) | 8.0** (15.5%) | 2.5 (4.8%) |
| Men | 62.4*** | 19.2*** (30.8%) | 9.9** (15.9%) | 14.6*** (23.4%) |
| Women | 40.8*** | 17.5*** (42.9%) | 6.6* (16.2%) | -13.4*** (-32.8%) |

*Statistically significant at the 10% level. ** Statistically significant at the 5% level. *** Statistically significant at the 1% level.

*Conclusion 4: For the pooled data, subjects with higher SAT scores and more agreeable types have significantly higher wage index functions, with a one standard deviation increase in SAT scores increasing this index function by a greater amount than a comparable increase in agreeableness. The regression estimates also indicate that men have significantly higher wage offer index functions than women. Dropping SAT from the regression leads to agreeableness becoming insignificant and extroversion becoming statistically significant, indicating the importance of including a measure of cognitive ability when assessing the impact of personality*

*characteristics on behavioral outcomes. Adding SAT and the Big Five considerably reduces the permanent unobserved heterogeneity in the model.*

Looking at separate estimates of the wage index function for men and women with SAT included (columns 3 and 4 of Table 2), the truly interesting result here is that conscientiousness is now statistically significant for both genders, but *opposite* in sign: It is positive for men but negative, with a comparable absolute value, for women. Although this opposite reaction by gender to increased conscientiousness is unexpected, some immediate justification for it can be found in observed differences in the degree of *reciprocity* resulting from increased conscientiousness found in the estimated effort response index functions reported below. There, other things equal, greater conscientiousness in men results in a modest but positive increase in the effort response (positive reciprocity), but in women it results in a modest *decrease* in the effort response. Although we did not anticipate different signs with respect to conscientiousness by gender, at least there is an internal consistency to these results, with more conscientious men, thinking from their own perspective, would be more likely to offer higher wages, while women thinking from their perspective would not.[35] The SAT coefficient value is comparable between men and women, as is agreeableness. Finally, dropping SAT (see columns 5 and 6) from these separate gender specifications again causes agreeableness to become insignificant. Adding the Big Five and SAT scores reduces the variance of the persistent individual unobserved heterogeneity by 31% and 19% for men and women, respectively, in the wage offer index function.[36] Finally, it is interesting to note that after controlling for SAT and the Big 5, the variance of the persistent individual heterogeneity for men is 50% larger than for women in the wage offer index function.

Rows (2) and (3) of Table 3 indicate the effect of a one standard deviation increase of key explanatory variables on the offered wage index function for men and women respectively. Here a change in conscientiousness results in a 14.6 ECU *increase* in the wage index function for men versus a 13.4 ECU *decrease* for women (a 23.4% increase compared to a 32.8% decrease from the mean of the index function). For men, this increase is almost as large as the impact of a

---

[35] A potential explanation for this differential effect is discussed in the conclusion.
[36] Recall that this is the reduction in the variance of the individual random effect.

one standard deviation increase in SAT scores (19.2 ECUs) and is larger than the impact of a one standard deviation increase in agreeableness (9.9 ECUs).

Conclusion 5: *Men and women differ substantially in their mean wage offer index functions, but show comparable effects in terms of a one standard deviation increase in SAT and agreeableness. However, the impact of conscientiousness is positive for men and negative for women, with both effects statistically significant and of comparable absolute value. This differential impact of conscientiousness is consistent with the estimated differential impact of conscientiousness on the effort response index functions for men and women reported on below. Adding SAT and the Big Five substantially reduces the permanent unobserved heterogeneity in the model for both men and women, with the permanent unobserved heterogeneity among men is 50% larger than that for women.*

.

*3.3 Statistical Analysis Including the Big Five and SAT: Effort Responses*

Since actual effort levels also are bounded by zero from below and 100 from above, we again use a random effects Tobit model for our statistical analysis. We assume that the index function for effort responses for individual $j$ in period $p$ whose manager is individual $i$ is given by

$$
\begin{aligned}
E_{jp}^* &= \delta_1 X_{jp} + \delta_2 w_{ip} + \delta_3 (w_{ip} * X_{jp}) + \gamma_j + \varepsilon_{jp} \\
&= \delta Z_{jp} + \gamma_j + v_{jp}.
\end{aligned}
\tag{3}
$$

Further, observed effort response is given by

$$
\begin{aligned}
E_{jp} &= 0 \quad \text{if} \quad E_{jp}^* < 0, \\
E_{jp} &= 100 \quad \text{if} \quad E_{jp}^* > 100, \\
E_{jp} &= E_{jp}^* \quad \text{otherwise.}
\end{aligned}
\tag{4}
$$

In (3) $\gamma_j$ is a random effects error term, which is *iid* across $j$ and distributed as $N(0, \sigma_\gamma^2)$, while $\varepsilon_{jp}$ is an idiosyncratic error term, which is *iid* (over $j$ and $p$) and distributed as $N(0, \sigma_\varepsilon^2)$.[37] Again

---

[37] We also assume that $\gamma_j$ and $u_{jp}$ are independent for all $j$ and $p$.

the variance of $\gamma_j$, $\sigma_\gamma^2$, represents a measure of the persistent unobserved subject heterogeneity in effort responses. We considered a version of (3) that contained session fixed effects, but this did not have any effect on the standard errors, and the session effects were not close to being significant at standard levels. We allow for interaction terms between the explanatory variables and the offered wage, since the null hypothesis of no interactions was decisively rejected (p < 0.01 in all cases).[38]

Table 4 reports the estimated effort response index functions in the same format as those for wage offer index functions. We restrict the analysis to the case where subjects face a positive wage offer, since zero wage offers are overwhelmingly met with zero effort. As such cognitive and non-cognitive characteristics essentially play no role in mediating responses to zero wage offers, so to include them would bias the estimates.[39] Column (1) of Table 4 presents the results for the pooled data when we include SAT.[40] The interaction effects complicate the interpretation of the overall effects of the independent variables. In what follows we start by evaluating effects at the mean values of the explanatory variables, but will extend the analysis beyond that as needed.[41] For example, at the mean values of SAT and the Big Five, a one-unit increase in the wage increases the effort response index function by 0.723; since the standard error associated with this responses is 0.033, this effect is strongly significant and precisely estimated.

The coefficient on the wage by SAT interaction term is positive and significant at the 10% level. This coefficient, in conjunction with the negative sign for the SAT coefficient itself, implies that subjects with higher SAT scores have a lower effort responses at all wages (other things equal), with this negative effect diminishing at higher wages.[42] At the mean wage rate the SAT effect (standard error) is -0.262 (0.324), so it is not statistically significant; it will be shown below that this occurs because SAT only has significant effects on male effort levels.

---

[38] The relationship between the index function $E^*$ and the expected value of effort is analogous to that between the index function $w^*$ and the expected value of the wage.

[39] The bias arises from mixing the effort index function (3) with the very different effort response index function that applies to a zero offered wage. Note that the offered wage is exogenous to the responders, hence omitting the responses for zero wage offers does not create any selection bias.

[40] For the pooled data, there are no significant ethnicity effects for any of the ethnicity dummies or for the wage by ethnicity interaction effects.

[41] In particular, responses to increases and decreases in the explanatory variables are no longer symmetric.

[42] $\partial E_{ip}^*/\partial SAT_i = -.601 + .0059 w_{ip}$, and only becomes positive for $w_{ip} > 101.9$, which is outside the [0,100] interval. A statistical test for SAT and SAT*wage together is not statistically significant at the 10% level or better. As will be shown below this results from the fact that the coefficients on SAT and SAT*wage are significant at conventional levels for men, but not for women, so that in the pooled data the combined effect is masked.

Table 4

Random Effects Estimates of the Effort Response Index Function

| Variable | Pooled data (men and women) | | Men | Women | Men | Women |
|---|---|---|---|---|---|---|
| Wage | 0.240 | 0.652** | -0.042 | 0.277 | 0.665 | 0.150 |
| | (0.380) | (0.306) | (0.621) | (0.579) | (0.557) | (0.417) |
| SAT | -0.601 | | -1.525*** | 0.097 | | |
| | (0.382) | | (0.550) | (0.527) | | |
| Agreeableness | 0.058 | 0.065 | 0.565 | 0.314 | 0.327 | 0.289 |
| (Agr) | (0.274) | (0.274) | (0.408) | (0.367) | (0.402) | (0.344) |
| Openness | 0.355 | 0.299 | 0.416 | -0.013 | 0.291 | 0.001 |
| (Ope) | (0.295) | (0.293) | (0.463) | (0.323) | (0.465) | (0.316) |
| Neuroticism | 0.107 | 0.076 | 0.319 | -0.146 | 0.242 | -0.140 |
| (Neu) | (0.221) | (0.220) | (0.324) | (0.269) | (0.326) | (0.269) |
| Conscientiousness | 0.641** | 0.631** | 0.770* | 0.230 | 0.769* | 0.234 |
| (Con) | (0.296) | (0.296) | (0.403) | (0.378) | (0.406) | (0.377) |
| Extroversion | -0.004 | 0.016 | -0.294 | 0.476* | -0.287 | 0.466* |
| (Ext) | (0.225) | (0.235) | (0.379) | (0.266) | (0.382) | (0.260) |
| Female | 5.961 | 7.765 | | | | |
| | (8.491) | (8.44) | | | | |
| Period | -0.405 | -0.422* | -0.516 | -0.157 | -0.558 | -0.153 |
| | (0.254) | (0.253) | (0.389) | (0.303) | (0.390) | (0.303) |
| Wage*SAT | 0.0059* | | 0.0134*** | -0.0145 | | |
| | (0.0032) | | (0.0050) | (0.0046) | | |
| Wage*Agr | 0.0122*** | 0.0121*** | 0.0087** | 0.0139*** | 0.0109*** | 0.0143*** |
| | (0.0023) | (0.0023) | (0.0036) | (0.0033) | (0.0035) | (0.0030) |
| Wage*Ope | -0.0035 | -0.0029 | -0.0002 | -0.0051* | 0.0012 | -0.0054* |
| | (0.0024) | (0.0024) | (0.0042) | (0.0028) | (0.0041) | (0.0028) |
| Wage*Neu | 0.0021 | 0.0025 | 0.0012 | 0.0060** | 0.0018 | 0.0059** |
| | (0.0018) | (0.0018) | (0.0030) | (0.0024) | (0.0029) | (0.0024) |
| Wage*Con | -0.0062*** | -0.0061** | -0.0092** | -0.0053* | -0.0088** | -0.0053* |
| | (0.0024) | (0.0024) | (0.0036) | (0.0031) | (0.0035) | (0.0031) |
| Wage*Ext | -0.0013 | -0.0015 | 0.0020 | -0.0028 | 0.0016 | -0.0026 |
| | (0.0018) | (0.0018) | (0.0033) | (0.0023) | (0.0033) | (0.0023) |
| Wage*Female | -0.227*** | -0.247*** | | | | |
| | (0.070) | (0.068) | | | | |
| Constant | -55.03 | -97.31*** | -59.27 | -64.36 | -139.1** | -56.22 |
| | (46.18) | (37.24) | (65.76) | (66.12) | (59.60) | (47.64) |
| Log Likelihood | -3303.7 | -3305.7 | -1821.3 | -1443.1 | -1826.0 | -1443.2 |

Standard Errors in Parentheses. *Statistically significant at the 10% level. ** Statistically significant at the 5% level.
*** Statistically significant at the 1% level.

The wage by agreeableness interaction term is positive and statistically significant at the 1% level, which, in conjunction with the positive coefficient for agreeableness, implies that more agreeable types offer higher effort at all wage rates. Specifically, at the mean wage rate the agreeableness effect (standard error) is a statistically significant 0.763 (0.236).    The coefficient for conscientiousness is positive and statistically significant, with the wage by conscientiousness coefficient negative and statistically significant. These coefficient values imply that more conscientious types provide higher effort throughout, with this positive effect diminishing at higher wages.  However, at the mean wage rate this effect (standard error) is 0.283 (0.252) and not statistically significant. Finally, none of the coefficients on extroversion, neuroticism and openness are statistically significant, with this being true for their overall effect as well.

Finally, the female by wage interaction effect is negative and statistically significant at the 1% level, indicating that women's effort responses are lower than for men.  This in conjunction with the small positive (but statistically insignificant) gender dummy, implies that in the pooled data, at lower wage rates, men and women have similar effort levels.  To be more precise, evaluated at the mean wage rate men provide higher effort on average than women, but the difference is not statistically significant (p = 0.29).  However, at higher wage rates the differences are statistically significant; e.g., evaluated at the 75th percentile for wages, p = 0.073.[43]

Since the overall SAT effect is small and insignificant in the pooled data, it is not surprising that dropping SAT does not affect the Big Five overall effects for the pooled effort data. A likelihood ratio test strongly test rejects the null hypothesis that all Big Five coefficients are zero, but dropping them does not affect the SAT coefficients. Finally adding the Big Five and SAT scores reduces the variance of persistent individual unobserved heterogeneity by 23% for the pooled effort response estimates.

Row (1) of Table 5 reports the quantitative effects of a one standard deviation increase in the key explanatory variables on the effort index function.  For the pooled data, wage plays the dominant role in determining effort levels, with a one standard deviation increase raising the effort index function by 22.4 ECUs (over 100%).  A one standard deviation increase in agreeableness increases the effort index function by 12.3 ECUs (65.1%).

---

[43] These wage percentiles based on wages greater than zero.

Table 5
Change in the Effort Index Function Resulting from One Standard Deviation
Increases in the Key Explanatory Variables
(Change as a Percent of the Mean Value of the Effort Index Function in Parentheses)

| | Mean Effort ($E$*) Index Function Value | Change in Mean Effort Index Function Value as a Result of a One Standard Deviation Increase | | | |
| --- | --- | --- | --- | --- | --- |
| | | Wage | SAT | Agreeableness | Conscientiousness |
| Pooled Data | 18.9*** | 22.4*** (118.5%) | -3.0 (-15.9%) | 12.3*** (65.1%) | 4.2 (22.5%) |
| Men | 19.8*** | 26.8*** (135.4%) | -8.9* (-44.9%) | 17.3* (87.4%) | 3.8 (19.2%) |
| Women | 16.9*** | 16.9*** (100.0%) | 0.1 (0.6%) | 18.0* (106.5%) | -1.1 (-6.5%) |

Effects are calculated at mean values for all variables. *Statistically significant at the 10% level. ** Statistically significant at the 5% level. *** Statistically significant at the 1% level.

Conclusion 6: *The estimated effort response index functions shows that overall, agreeable types provide higher effort, but that the overall effects for SAT,  conscientious and gender are not statistically significant at the mean gift level. On average, a unit increase in wages is met with a 0.75 increase in effort. A one standard deviation increase in agreeableness has a large impact (a 65% increase) on the effort response function, but a one standard deviation increase in wages has an even larger impact. Men provide significantly higher effort levels than women at higher wage rates. Adding SAT and the Big Five substantially reduces the permanent unobserved heterogeneity in the model.*

The remaining columns in Table 4 report separate estimates of the effort response index function for men and women, since a likelihood ratio test decisively rejects the null hypothesis

that men and women have the same coefficients except for the intercept.[44] Columns (3) and (4) are our preferred specification where SAT is included. For men and women the effort responses (standard error) to the wage, at the mean values of all of the other explanatory variables, are 0.866 (.051) and 0.550 (.039), respectively, so that men **p**rovide significantly higher effort *responses* at the mean wage rate.[45]

Note that statistically significant coefficients for SAT are confined exclusively to men, where the coefficient on the SAT variable by itself is negative and significant at the 1% level, with the wage by SAT variable coefficient positive and significant ($p < 0.01$) as well. The estimates indicate that men with higher SAT scores provide lower effort throughout the range of possible wages, with its major impact confined to lower wage offers. At the mean wage rate, the overall SAT effect for men (standard error) is -0.757 (.444), [46] which is significant at the 10% level, and implies that a one standard deviation increase in SAT *reduces* the male effort index function by 44.9%. In contrast, for women, neither of the two SAT variables is individually significant, nor are they jointly significant at conventional test levels.[47] Moreover, at the mean wage rate, the overall SAT response (standard error) for women is a tiny 0.013 (.447). One possible interpretation of the negative male SAT effect is that men with greater cognitive ability are more sensitive to the one-shot nature of the interactions inherent in the structure of the game. It is not clear why women would be less sensitive in this dimension given that they have comparable SAT scores.[48] The fact that the negative SAT effect is diminished at higher wage rates suggests mixed motives at work for higher SAT type men: they remain more reciprocal at higher wages, but offer lower effort holding the wage constant, than lower SAT men or all women, perhaps on account of the one-shot nature of the game. Nevertheless, it is considerably more profitable, on average, to offer a higher wage to a high SAT man, than an average wage to an average SAT man (124 ECUs versus 42 ECUs).[49]

---

[44] For example, when we include SAT the test statistic is $78.6 > 37.57 = \chi^2_{99\%}(20)$.

[45] In what follows we use the men's mean for their effects and the women's mean for the women's effects. However, evaluating all effects at the pooled mean, men's mean, or women's mean has only a trivial impact on the overall effects.

[46] For example, the effect of a one-unit increase in SAT at wage rate 85.0, is -0.384, which is considerably smaller than the effect of -0.757 obtained at the mean wage of 57.2 reported in the text.

[47] The test statistic is 0.10 with a p-value of 0.951.

[48] This result is consistent with the evidence that it takes women longer to learn to act strategically in some game theoretic settings than men (see Casari, et al., 2007 and Cooper and Kagel, 2012b).

[49] This is after subtracting out the 100 ECUs in the manager's payoff function: High wage is one standard deviation above the average wage, high SAT is one standard deviation above the average SAT.

The agreeableness coefficients are positive but not statistically significant for men and women, but the wage by agreeableness coefficients are statistically significant in both cases. The overall effect of agreeableness (standard errors) at the mean wage rate is to increase the effort response by 1.06 (.337) and 1.12 (.316) for men and women respectively. Note that these are effects very similar for men and women and both are statistically significant, with the respective entries in Table 5 reporting that a one standard deviation in agreeableness leads to an 87.4% and 106.5% increase in the effort response index function respectively for men and women. It is quite impressive that one of the personality characteristics – agreeableness – can have such a large quantitative impact on the effort response index function compared to the core economic variable wage paid, a result that is inconsistent with purely selfish economic man, but is consistent with the growing economics literature on other regarding preferences.

The coefficient for conscientiousness by itself is positive for both men and women, but for men the value is considerably larger and statistically significant. The coefficient on the conscientiousness by wage variable is significantly negative for both men and women, but again the effect for men is bigger (in absolute value). At the mean wage rate, the overall impact of conscientiousness (standard error) is to raise the effort response of men by 0.245 (.329) while for women there is a slight reduction, -0.078 (.324) with neither effect being statistically significant at conventional levels. For men, at lower average wages, an increase in conscientiousness results in a statistically significant increase in effort levels.[50] At higher wages more conscientious women have significantly lower effort levels.[51] Focusing on the signs of the conscientiousness effects in Tables 3 and 5, this differential effort response by men and women with respect to conscientiousness is internally consistent with the effect this variable has on male and female wage offers. That is, both genders may accurately predict own gender effort response with respect to conscientiousness, and act accordingly in setting wages, even though they do not know the gender of the person they are interacting with in any given play of the game. Note there is independent evidence for this sort of effect in the psychology literature, referred to as "consensus bias" - the overuse of self-related knowledge in estimating the prevalence of attributes in a population (Ross, Green and House, 1977; Kruger and Clement, 1994).

---

[50] A 1.5 standard deviation decrease in wages evaluated at the average level of conscientiousness (p = 0.08).
[51] A 1 standard deviation increase in wages in conjunction with a 1 standard deviation increase in conscientiousness (p = 0.01).

In terms of neuroticism, the women's (positive) wage by interaction is statistically significant but the overall effects are relatively small and insignificant for both genders at the mean wage rate. Women have a significantly positive extroversion coefficient but again the overall effects are relatively small and insignificant for both genders at the mean wage rate. Finally, the coefficient on the openness-wage interaction is negative and significant at the 10% level, but the other openness coefficient and the overall openness effect are statistically insignificant for both men and women.

Finally, when we analyze men and women separately, we find that adding the Big Five and SAT scores reduces the variance of the persistent individual unobserved heterogeneity by 33% and 35% for men and women respectively. In addition, the men's variance of the persistent individual unobserved heterogeneity is twice as large as it is for women.

Unlike the wage offer index function, dropping SAT does not have an important effect on the Big Five variables for any of our samples. However, it is interesting to note that a second order effect of dropping SAT is found in the race by wage interaction effects (not reported in the table but included in the specifications). With SAT in, the only statistically significant race effect is that the dummy variable for Asian men is positive and significant at the 10% level. With SAT out, the wage by African-American interaction effect is negative and statistically significant at the 5% level for men. (With SAT in, the African-American dummy variable is not significant at conventional levels.) These ethnicity effects suggest that it will be important to control for SAT in any study examining ethnic (as opposed to gender) differences in experiments.

Conclusion 7: *There are major differences in effort responses of men and women: Men with higher SAT scores offer less effort than those with lower SAT scores, particularly at the low end of the wage scale, but SAT scores have minimal impact on the effort provided by women. More conscientious men supply greater effort, particularly at lower wages, while greater conscientiousness among women results in less effort at higher wages. A one standard deviation increase in average wages increases the effort response index function for men by 135% compared to 100% for women, with a one standard deviation in agreeableness having substantial effects for both men and women as well. Dropping SAT does not affect the Big Five coefficient estimates but has a substantial impact on the ethnicity coefficients. Adding SAT and the Big Five reduces the permanent unobserved heterogeneity in the effort response index*

*function substantially for both men and women, with the permanent unobserved heterogeneity being 100% larger for men than for women.*

### 4. Summary and Conclusions

We report results from a one-shot gift exchange experiment accounting for the effects of gender, cognitive ability, and the Big Five personality characteristics on outcomes. We find substantial impacts on behavior for each of these typically neglected factors.[52] On average, women offer lower wages than men do when they are evaluated at the same level of the explanatory variables, with women offering less effort than men in response to the same wage offers. These results add to the growing literature on gender effects in economics and in particular on gender effects in economic experiments (see Croson and Gneezy, 2009 for a survey). As in any experiment, there is the danger that gender differences simply reflect very different men and women volunteering for the experiment. To investigate this we conduct balancing tests, concluding that the gender differences reported do not simply reflect differential selection effects.

The major impact of cognitive ability on outcomes is that both men and women with higher SAT scores offer higher wages than their counterparts with lower SAT scores. We conjecture that the primary factor behind this outcome is that higher cognitive ability types are better attuned to large potential profits associated with higher wages (Figure 3), and better able to tolerate the risk associated with offering these higher wages (Dohmen et al., 2010; Burks et al., 2009). Dropping SAT from the Tobit regressions for wages has several effects: For the pooled data, the coefficient value for agreeableness goes from being positive and statistically significant at better than the 5% level, with SAT in, to no longer being statistically significant. Further extroversion becomes statistically significant at the 1% level with SAT out, indicating the importance of having a measure of cognitive ability when investigating the impact of personality characteristics on economic behavior. Dropping SAT from the effort equation also affects the size and significance of ethnicity effects, indicating the potential importance of including some measure of cognitive ability when investigating ethnicity differences in effort responses.

---

[52] In turn these hidden personality and cognitive ability factors may help account for the high degree of individual subject variability in effort responses and wage offers reported in the typical gift exchange experiment.

At times, the Big Five personality characteristics have as large an impact on the wage offer and effort response index functions as cognitive ability and economic variables (wages). As in most experiments of this sort, higher wage offers are met with a higher effort response. A one standard deviation increase in agreeableness has, for women, the same impact on the effort response index function as does a one standard deviation increase in wages, while having an impact on male effort just under two-thirds the impact of a comparable wage increase. On the wage side, for men, the impact of a one standard deviation increase in conscientiousness increases the wage offer index function by about the same amount as a one standard deviation increase in SAT scores.

The surprising result here is that, for women, conscientiousness has the *opposite* impact on the wage offer index function (of roughly the same absolute value) as it does for men. This differential effect of conscientiousness on wages is consistent with best responding to its effect on effort where, at low wages, increased conscientiousness leads to increases in the effort response index for men but essentially the same or a modest negative effect for women. One possible explanation for this differential effect of conscientiousness is as follows: One element of the conceptual definition of conscientiousness is "following norms and rules" (John et al., 2008, Table 4.2). With this in mind, note that there is some evidence suggesting that, for women, explicit monetary payments tend to drive out social preferences more than for men.[53] In this case, more conscientious women would be more likely to have lower responsiveness to wage offers, with women wage givers best responding to these beliefs. In contrast, if men are less sensitive, or immune, to this crowding out effect, and more accepting of the notion of explicit monetary benefits for reciprocal responses, more conscientious men would be more likely to take account of this fact and offer higher wages.

Finally, adding the Big Five and SAT significantly reduces the permanent unobserved heterogeneity in the pooled sample, as well as for men and women for both the wage offer index function and the corresponding effort response index function. Quite unexpectedly, the variance for the men's unobserved permanent error component is 50% and 100% larger than for women in the wage offer and effort response index functions, respectively.

---

[53] Mellström and Johannesson (2008) found that paying people to donate blood reduced women's donations, while men's donations were unaffected. However, others have failed to find any evidence of gender effects in blood donations in large-scale natural field experiments (see Lacetera, Macis, and Slonim, 2011).

The results of this experiment have obvious and immediate implications for the social preference literature in economics. We believe that they have implications for the labor economics literature as well. On this last point, one of the interesting questions will be to extend the analysis of the role of the Big Five personality characteristics, gender, and cognitive ability to gift exchange games in which agents can develop reputations through repeated or longer term contracts. In this case, we would expect an even bigger impact of cognitive ability on effort responses for both men and women, with agents with greater cognitive ability providing greater effort in response to higher wages, motivated by the potential for cooperation inherent in repeated interactions.

# References

Almlund, Mathilde, Angela Lee Duckworth, James J. Heckman, and Tim Kautz, 2011. ìPersonality Psychology and Economics" in *Handbook of the Economics of Education Vol. 4,* E. A. Hanushek, S. Machin, and L. Woessman (Eds.), North Holland:Elsevier: 1-181.

Anderson, Jon, Stephen Burks, Colin De Young, and Aldo Rustichini, 2011. "Toward the Integration of Personality Theory and Decision Theory in Explanation of Economic Behavior," *University of Minnesota working paper*.

Babcock, R. 1994. "Analysis of Adult Age Differences on the Raven's Advanced Progressive Matrices Test." *Psychology and Aging,* Vol. 9, pp. 303-314.

Becker, Anke, Thomas Deckers, Thomas Dohmen, Armin Falk, and Fabian Kosse, 2012. "The Relationship Between Economic Preferences and Psychological Personality Measures," working paper, University of Bonn.

Borghans, L., A.L. Duckworth, J.J. Heckman, and B.T. Weel, 2008. "The Economics and Psychology of Personality Traits," *Journal of Human Resources*, Vol. 43, pp. 972-1059.

Bors, D. and F. Vigneaub, 2001. "The Effect of Practice on Raven's Advanced Progressive Matrices." *Learning and Individual Differences,* Vol. 13, pp. 291–312.

Brouwers , S., F. J.R. Van de Vijver, and D. Van Hemert, 2009. "Variation in Raven's Progressive Matrices Scores Across Time and Place." *Learning and Individual Differences,* Vol. 19, pp. 330–338.

Burks, S.V., J. Carpenter, L. Götte, A. Rustichini, 2009. "Cognitive Skills Affect Economic Preferences, Social Awareness, and Job Attachment," *Proceedings of the National Academy of Science*, Vol. 106, pp. 7745-7750.

Casari, Marco, Ham, John, and Kagel, John H., 2007. "Selection Bias, Demographic Effects and Ability Effects in Common Value Auction Experiments," *American Economic Review*, vol. 97 (4), pp. 1278-1304.

Cohen, Patricia,, Jacob Cohen, Leona S. Aiken and Stephen G. West, 1999. "The Problem of Units and the Circumstance for POMP," *Multivariate Behavioral Research*, Vol. 34, pp. 315-346.

Cooper, David J. and John H. Kagel, 2012a. "Other Regarding Preferences: A Selective Survey of Experimental Results," to appear in *The Handbook of Experimental Economics*, *vol.* 2, John H. Kagel and Alvin E. Roth (Eds), Princeton University Press.

Cooper, David J. and John H. Kagel, 2012b. "A Failure to Communicate: An Experimental Investigation of the Effects of Advice on Strategic Play," Ohio State University working paper.

Croson, Rachel and Uri Gneezy, 2009. "Gender Differences in Preferences," *Journal of Economic Literature*, Vol. 47, pp. 448-474.

Dohmen, Thomas and Armin Falk, 2011. "Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender," *American Economic Review*, Vol. 101, pp. 556-590.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde, 2010. "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review*, Vol. 100, pp. 1238-1260.

Duffy, John and Jack Ochs, 2009. "Cooperative Behavior and the Frequency of Social Interaction," *Games and Economic Behavior,* Vol. 66 (2), pp. 785-812.

Eckel, Catherine C., and Rick K. Wilson, 2004. "Is Trust a Risky Decision?" *Journal of Economic Behavior and Organization*, Vol. 55 (4), pp. 447-65.

Einav, Liran, Amy Finkelstein, Iulian Pascu and Mark R. Cullen, 2012. "How General Are Risk Preferences? Choices Under Uncertainty in Different Domains," *American Economic Review*, Vol. 102(6), pp. 2606-2638.

Fischbacher, Urs 2007. "z-Tree: Zurich Toolbox for Ready Made Economic Experiments," *Experimental Economics*, Vol. 10, pp. 171-178.

Frey, Meredith C. and Douglas K. Detterman, 2004. "Scholastic Assessment or *g*?: The Relationship Between the Scholastic Assessment Test and General Cognition Ability," *Psychological Science*, Vol. 15 (6),  pp. 373- 378.

Gächter, Simon and Ernst Fehr, 2001. "Fairness in the Labour Market –A Survey of Experimental Results" http://ideas.repec.org/p/zur/iewwpx/114.html

Heckman, James J., Jora Stixrud, and Sergio Urzua, 2006.  "The Effects of Cognitive and Noncognitive Abilities on Labor      Market Outcomes and Social Behavior," *Journal of Labor Economics*, Vol. 24 (3),  pp. 411 - 482.

Imbens, Guido and Jeffrey Wooldridge, 2009.  *Lecture Notes for Applied Microeconometrics Workshop*, Institute for Research on Poverty, University of Wisconsin, Madison, August 2009. http://www.irp.wisc.edu/newsevents/workshops/appliedmicroeconometrics/schedule1.htm

John, Oliver P., Naumann, Laura, P., and Christopher J. Soto, 2008. "Paradigm Shift: to the Integrative Big Five." In Oliver P. John, R. W. Robbins, and L. A. Pervin (Eds.), *Handbook of Personality, Theory, and*  Research, N. Y. Guilford Press, pp. 114-158.

Jones, Matthew T. 2012. "Strategic Complexity and Cooperation: An Experimental Study," SSRN Working Paper. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2025659

Kandori, Michihiro 1992. "Social Norms and Community Enforcement," *Review of Economic Studies,* Vol. 59, pp. 63–80.

Kern, M., A. Duckworth, S. Urzua, R. Loeber, M. Loeber, and D. Lynam,  2013. ."Mind Your Manners! Aspects of Agreeableness and Early Adult Outcomes for Inner---City Boys." Mimeo, University of Pennsylvania.

Kruger, Joachim and Russell W. Clement, 1994. "The Truly False Consensus Effect: An Ineradicable and Egocentric Bias in Social Perception," *Journal of Personality and Social Psychology*, Vol. 67 (4), pp. 596-610.

Kurzban, Robert and Daniel Houser, 2001. "Individual Differences in Cooperation in a Circular Public Goods Game," *European Journal of Personality*, Vol. 15, pp. 37-52.

Lacetera, N., Macis, M., and R. Slonim, 2011. Rewarding altruism: A natural field experiment. NBER Working Paper 17636.

Lazear, E. and P. Oyer, 2013. "Personnel Economics." In: Robert Gibbons and D. John Roberts, *The Handbook of Organizational Economics,* Princeton University Press, Princeton NJ.

Mellström, Carl and Magnus Johannesson, 2008. "Crowding Out in Blood Donations: Was Titmuss Right?" *Journal of the European Economic Association*, Vol. 6, pp. 845-863.

Pothos, Emmanuel M., Gavin Perry, Philip J. Corr, Mervin R. Matthew, and Jerome R. Busemeyer, 2010. "Understanding Cooperation in the Prisoner's Dilemma Game," *Personality and Individual Differences*, Vol. 51, pp. 210-216.

Ross, Lee, Greene, David, and Pamela House, 1977. "The "False Consensus Effect": An Egocentric Bias in Social Perception," *Journal of Experimental Social Psychology,* Vol. 13, pp. 279-301.

Skuya, M., A. Gewera, Y. Osrina, D. Khunoua and P. Fridjhona, 2002. "Effects of Mediated Learning Experience on Raven's Matrices Scores of African and non-African University Students in South Africa." *Intelligence,* Vol. 30, pp. 221 – 232.

Urzua, S. and Gregory Veramendi, 2012. "Empirical Strategies to Identify the Determinants And Consequences of Skills," Mimeo, University of Maryland.

**Review of the Theil-Griliches OLS Specification Error Result for Omitting SAT**

**In the Wage Offer Equations**

Since we have found that some experimenters are unfamiliar with the Theil-Griliches (TG) specification error result, we review it briefly here. We use it to understand the biases from omitting SAT in the wage offer equation; because it applies to linear regressions while we estimate a two limit, random effects, Tobit model, this discussion is illustrative. Assume the true wage offer linear equations is given by

$$w_{ip} = X_{1ip}\beta_1 + SAT_i\beta_2 + e_{ip}. \tag{A1}$$

However, we omit $SAT_i$ from (A1) and estimate instead

$$w_{ip} = X_{1ip}\gamma + u_{ip}, \qquad \text{where } u_{ip} = SAT_i\beta_2 + e_{ip}. \tag{A2}$$

Define the projection (linear regression)[54]

$$SAT_{ip} = X_{1ip}\pi + v_{ip}, \quad i = 1,...N, \ p = 1,...,P. \tag{A3}$$

For an independent variable $x_{kip}$ in $X_{1ip}$, the TG result implies that

$$\hat{\gamma}_k = \hat{\beta}_1 + \hat{\beta}_2\hat{\pi}_k \quad \text{and} \quad \text{plim}(\hat{\gamma}_k) = \beta_1 + \beta_2\pi_k. \tag{A4}$$

Thus if $\beta_2 > 0$ (as column (1) of Table 2 strongly suggests for the SAT coefficient), then its coefficient $\hat{\gamma}_k$ in (A2) will be biased upwards if its coefficient $\pi_k$ in a multiple regression of $SAT_i$ on $X_{ip}$ is (asymptotically) positive. The following table presents the estimates of $\pi$ from (A3).

---

[54] Thus for individual $i$ $\ SAT_{ip} = SAT_i$ for all periods $p$.

## Table A

### Regression of SAT on the Other Independent Variables

| Variable | Pooled data (men and women) | Men | Women |
|---|---|---|---|
| Agreeableness (Agr) | -0.0701 (0.0504) | -0.0482 (0.0743) | -0.175*** (0.0616) |
| Openness (Ope) | 0.0915* (0.0525) | 0.136 (0.0835) | 0.0663 (0.0572) |
| Neuroticism (Neu) | -0.00901 (0.0443) | -0.0349 (0.0682) | -0.0355 (0.0504) |
| Conscientiousness (Con) | 0.0283 (0.0597) | 0.104 (0.0859) | -0.0511 (0.0735) |
| Extraversion (Ext) | -0.142*** (0.0443) | -0.129* (0.0713) | -0.163*** (0.0472) |
| Female | -1.706 (1.611) | | |
| White | 3.847 (2.957) | 4.416 (4.282) | 4.676 (3.501) |
| Black | -8.312** (3.328) | -2.778 (5.108) | -13.89*** (3.77) |
| Asian | 1.402 (3.13) | 5.667 (4.734) | -3.158 (3.571) |
| Constant | 78.03*** (7.077) | 66.83*** (10.49) | 95.09*** (8.486) |
| Observations | 192 | 105 | 87 |

Standard errors in parentheses. *Statistically significant at the 10% level. ** Statistically significant at the 5% level. *** Statistically significant at the 1% level.